Simulating the replication of some of Greenberg's word order generalizations

by MARTIN HASPELMATH and SVEN SIEGMUND

Scientific generalizations must be reproducible (or replicable) to be accepted as valid, whether they are based on simple observations or controlled observations, i.e., experiments. In linguistics, issues of replicability arise at various levels:

(i) replicability of speaker behavior under identical experimental conditions (acceptability judgments, elicitation of narratives);
(ii) replicability of descriptions of the same language (either the same corpus: will two linguists arrive at the same description of Gothic on the basis of the Wulfila corpus?; or the same speech community: will two fieldworkers doing fieldwork in the same village arrive at the same grammars?);
(iii) replicability of typological coding based on identical descriptions (will two typologists extract the same information from the same grammars?);
(iv) replicability of typological generalizations based on different samples (drawn from the same universe of languages).

At all these levels, research results should in principle be replicable. For linguistic typology, (iii) and (iv) are the most relevant. In addition to replicating typological generalizations based on a different sample, one could also ask for typological replication based on the same sample (as suggested by a reviewer). But the issues arising in such an exercise would not be distinct from (ii) and (iii). We feel that (iv) is a much more interesting issue for typologists, and we restrict our attention to it in this note.[1]

Full-scale replication of an earlier researcher's results is rarely done in linguistics, perhaps because linguists value novelty more than reliability: just confirming observations made by others is simply not particularly prestigious. Fieldworkers don't generally write another grammar or another dictionary of a language spoken by a small community if a comprehensive grammar and a large dictionary have been published within the last decade. With so many languages disappearing, we cannot afford that.

In language typology, the situation is similar: typologists do not seem to regard the replication of known facts as sufficiently prestigious to be worth the effort. This is so despite the fact that the empirical basis of typological generalizations is often far from ideal: language samples tend to be biased; it is even unclear what an unbiased sample would be; grammars show widely varying degrees of thoroughness and explicitness; they adopt different theoretical

1. A short discussion of (iii) ("inter-coder reliability") is found in Bybee, Perkins, & Pagliuca (1994: xvii).

outlooks; and so on. Every practitioner knows that doing typological surveys is not easy, and some critics have voiced skepticism about the entire typological enterprise as a result of these difficulties (Newmeyer 1998: Ch. 6).

Thus, we feel that the field of typology would be strengthened as a whole if it could be shown that typological generalizations can generally be successfully replicated on a different sample. A typological survey is basically like an opinion poll, and opinion polls are routinely replicated by competing polling institutions, who ask a different sample of people. Another analogy is that of a corpus study, which can be replicated by looking at the same kinds of phenomena in a different corpus.[2]

In this short contribution, we want to illustrate what a typological replication in this sense would look like by examining some of Greenberg's (1963) word order generalizations in a range of different 30-language samples. This is not a real replication based on new data (which would have required a substantial amount of research, beyond the scope of this discussion note), but a "simulated" replication in that it is based on data gathered and published independently: Matthew Dryer's word order data published as individual chapters of the *World Atlas of Language Structures* (*WALS*; Dryer 2005a–e).[3] What we did for this note was to extract data for a range of different 30-language samples from Dryer's large database, in order to compare them with Greenberg's original findings. In Dryer's data, there are 656 languages that had complete information for our purposes. Our 30-language samples are drawn from this large set of languages.

Of course, word order typology is the one area within typology where successful replicability has already been demonstrated. The studies of Hawkins (1983) and various earlier publications by Matthew Dryer (especially Dryer 1992) have looked at many of the same typological parameters as Greenberg, using essentially the same criteria, but substantially expanded samples (which is why they have been regarded as worthwhile). And by and large, Greenberg's results have been confirmed by these later studies, although counterexamples were found to most of the universals that seemed exceptionless to Greenberg, and it turned out that sampling bias could in some cases lead typologists seriously astray (see especially Dryer 1988).

---

2. A nice example of a well-known corpus study that has been widely replicated is Du Bois's Preferred Argument Structure generalizations, originally based on data from Sakapultek. Du Bois claimed that the observed pattern should be universal, and in fact his study has been replicated numerous times for different languages (see Du Bois et al. (eds.) 2003 and the references cited there). Here the fact that different languages were involved helped, because doing the same study for a new language was apparently considered sufficiently prestigious.
3. Since the typological data were gathered by a different researcher, the replication also involves issue (iii), and since Dryer's sources were presumably sometimes different from Greenberg's, issue (ii) also arises. We assume here that their effects can be neglected in the present context.

Table 1. *Greenberg's (1963) sample*

1a.	VSO	SVO	SOV	other
Postp	0	3	11	0
Prep	6	10	0	0
other	0	0	0	0

1c.	NAdj	AdjN	other
DemN	12	7	0
NDem	11	0	0
other	0	0	0

1b.	VSO	SVO	SOV	other
AdjN	0	5	6	0
NAdj	6	8	5	0
other	0	0	0	0

1d.	NAdj	AdjN	other
NumN	8	10	0
NNum	11	0	0
other	0	0	0

Sample languages:

Africa	7 languages: Berber, Fulani, Masai, Nubian, Songhai, Swahili, Yoruba
Australia-New Guinea	1 language: Loritja
Eurasia	13 languages: Basque, Burushaski, Finnish, Greek, Hebrew, Hindi, Italian, Japanese, Kannada, Norwegian, Serbian, Turkish, Welsh
North America	2 languages: Maya, Zapotec
SE Asia and Oceania	4 languages: Burmese, Malay, Maori, Thai
South America	3 languages: Chibcha, Guaraní, Quechua

Here we will look at just five word order parameters: the order of subject, object, and verb (Dryer 2005a), the order of adposition and noun phrase (Dryer 2005b), the order of adjective and noun (Dryer 2005c), the order of demonstrative and noun (Dryer 2005d), and the order of numeral and noun (Dryer 2005e). We will present the results in four tables for each sample, corresponding to Greenberg's (1963) Tables 1, 5, and 6.

Let us first look at Greenberg's 30-language sample in Table 1. Greenberg derived the following six universals from these figures:

Universal 1:	In declarative sentences with nominal subject and object, the dominant order is almost always one in which the subject precedes the object.
Universal 3:	Languages with dominant VSO order are always prepositional.
Universal 4:	With overwhelmingly greater than chance frequency, languages with normal SOV order are postpositional.
Universal 17:	With overwhelmingly more than chance frequency, languages with dominant order VSO have the adjective after the noun.
Universal 18:	When the descriptive adjective precedes the noun, the demonstrative (18a) and the numeral (18b), with overwhelmingly more than chance frequency, do likewise.

Table 2. *A "West Papuan" sample*

2a.	VSO	SVO	SOV	other
Postp	0	0	14	4
Prep	1	8	0	2
other	0	1	0	0

2c.	NAdj	AdjN	other
DemN	6	7	2
NDem	14	0	0
other	0	0	1

2b.	VSO	SVO	SOV	other
AdjN	0	1	4	2
NAdj	1	8	8	3
other	0	0	2	1

2d.	NAdj	AdjN	other
NumN	4	6	2
NNum	15	0	1
other	1	1	0

Sample languages:

Africa	1 language: Moro
Australia-New Guinea	12 languages: Amele, Arrernte (Mparntwe), Bilua, Dera, Hatam, Koiali (Mountain), Kombai, Mpur, Ndjébbana, Tabla, Una, Wardaman
Eurasia	4 languages: Italian, Mansi, Nivkh, Tulu
North America	3 languages: Biloxi, Guarijío, Tlingit
SE Asia and Oceania	8 languages: Ambai, Nguna, Nias, Pattani, Tetun, Timugon, Palauan, Rukai
South America	2 languages: Epena Pedee, Wichí

Universal 1 corresponds to the zeroes in the last column of Table 1a, and each of the other universals corresponds to one of the zeroes in the boxed areas.

To replicate this result, we need to look at a different sample of 30 languages. There are various ways to do this, and sampling techniques have been widely discussed in the typological literature. Here we will start by considering two samples that are not better than Greenberg's sample, which was heavily biased toward Eurasian languages. In Greenberg's sample, 13 languages are from Eurasia, 7 of them from Europe. Africa is represented by 7 languages, Southeast Asia (with Oceania) by 4, Australia and New Guinea by 1, North America by 2 (both from Central America), and South America by 3.

Our first biased sample, which might have been constructed by a West Papuan linguist with little knowledge of distant continents, contains 12 languages from Australia and New Guinea, 7 of them from West Papua (in Indonesia, formerly Irian Jaya). Southeast Asia is represented by 8 languages, Eurasia by 4, and so on (see Table 2). As in Greenberg's sample, each language here is from a different genus. (Greenberg did not explicitly say this, and the term was of course only introduced by Dryer 1989, but Greenberg clearly avoided choosing two languages from a close-knit grouping that we now call a genus.)

Perhaps surprisingly, we see that the results are quite similar to Greenberg's. Within the boxes, all of Greenberg's zeroes correspond to zeroes here, so a

Table 3. *A "Brazilian" sample*

3a.	VSO	SVO	SOV	other
Postp	1	1	15	5
Prep	2	3	0	2
other	0	0	1	0

3c.	NAdj	AdjN	other
DemN	12	8	1
NDem	8	1	0
other	0	0	0

3b.	VSO	SVO	SOV	other
AdjN	0	1	5	3
NAdj	3	3	10	4
other	0	0	1	0

3d.	NAdj	AdjN	other
NumN	11	8	1
NNum	9	1	0
other	0	0	0

Sample languages:

Africa	3 languages: Korana, Nzakara, Tubu
Australia-New Guinea	2 languages: Ngaanyatjarra, Sulka
Eurasia	4 languages: Bhojpuri, Buriat, Itelmen, Tsova-Tush
North America	7 languages: Chinantec (Palantla), Cora, Koasati, Mandan, Nevome, Sarcee, Zuni
SE Asia and Oceania	1 language: Yapese
South America	13 languages: Cayuvava, Dâw, Guaraní, Kariri, Macushi, Mapudungun, Ngäbere, Palikur, Paumarí, Pirahã, Shiriana, Tsafiki, Warao

West Papuan typologist could have discovered more or less the same universals. (The languages corresponding to the figures outside the boxes, in the columns and rows labeled "other", mostly have no dominant word order, a possibility that Greenberg simply ignored.) Since there is an additional zero in Table 2a, indicating the absence of postpositional SVO languages, Universal 4 could be strengthened to "VSO or SVO → Prep". However, the West Papuan typologist would probably have refrained from making any generalizations about VSO languages, given that there is only one such language in the sample. Greenberg's Universal 1 is also largely confirmed, though while it was exceptionless in Greenberg's sample, there is one VOS language in this sample (Nias, an Austronesian language of the Sumatra area).

What if our replication effort is undertaken in Brazil, and all we can achieve is a biased sample with 13 South American languages (7 of them from Brazil), 7 North American languages, and so on? The results are shown in Table 3.

Here both Tables 3c and 3d have an isolated counterexample to universals 18a/b (Nzakara, in Central Africa), but otherwise the picture is not worse. This sample has three VSO languages, and they all conform to Universal 17 (i.e., they are all NAdj). However, there is a counterexample to Universal 3 (one postpositional VSO language: Cora, a Uto-Aztecan language), so we would not be too confident about this one. VSO languages would perhaps not have been

Table 4. *A random sample*

	VSO	SVO	SOV	other
Postp	0	2	10	1
Prep	3	8	1	2
other	0	0	3	0

	NAdj	AdjN	other
DemN	6	8	2
NDem	10	2	1
other	1	0	0

	VSO	SVO	SOV	other
AdjN	0	4	5	2
NAdj	3	5	7	1
other	0	1	2	0

	NAdj	AdjN	other
NumN	7	8	0
NNum	9	2	3
other	1	0	0

Sample languages:

Africa	10 languages: Amharic, Arabic (Modern Standard), Basaá, Fula (Cameroonian), Igbo, Kunama, Linda, Nzakara, Obolo, Somali
Australia-New Guinea	4 languages: Barai, Edolo, Ndjébbana, Yidiny
Eurasia	6 languages: Estonian, Karakalpak, Persian, Turkmen, Ubykh, Wakhi
North-America	2 languages: Chatino (Yaitepec), Sarcee
SE Asia and Oceania	6 languages: Cantonese, Chang, Hawaiian, Kachari, Leti, Nias
South America	2 languages: Apurinã, Cayuvava

prominent for the Brazilian typologist anyway, because this sample also has two VOS languages (Cayuvava, in Bolivia, and Kariri, in Brazil), as well as one OSV language (Warao, in Venezuela). Thus, there are three counterexamples to Universal 1 here, but the SO/OS asymmetry is still striking.

Now suppose that we want to avoid regional bias by selecting 30 languages completely randomly from all over the world. One such random selection, using the randomized sequence generator at the website random.org, gives us Table 4. The random selection here is of course actually from those 656 languages for which Dryer's *WALS* chapters have all the relevant data.

Again, the picture that we get is fairly similar to the pictures of Tables 1 to 3. There is a little more support for the universals involving VSO languages, and one counterexample to Universal 4 (the notorious prepositional SOV language Persian). However, we have two counterexamples to Universals 18a/b, which could be enough to shake our belief in them. But when we look at these two languages more closely, we see that they are closely related: Linda and Nzakara, two languages of Central Africa belonging to the same genus, Adamawa-Ubangian. Unlike Greenberg's sample, this sample did not try to avoid genealogical bias in this way, and as a result we have four pairs of languages from the same genus (Linda/Nzakara, Amharic/Arabic, Persian/Wakhi,

Table 5. *An areally and genealogically (by genus) stratified sample*

	VSO	SVO	SOV	other
Postp	0	1	11	3
Prep	2	6	1	4
other	0	2	0	0

	NAdj	AdjN	other
DemN	9	11	0
NDem	8	1	1
other	0	0	0

	VSO	SVO	SOV	other
AdjN	1	4	3	4
NAdj	1	5	8	3
other	0	0	1	0

	NAdj	AdjN	other
NumN	5	12	0
NNum	11	0	1
other	1	0	0

Sample languages:

Africa	5 languages: Gula (in Central African Republic), Iraqw, Moro, Ngiti, Ogbronuagum
Australia-New Guinea	5 languages: Hamtai, Kire, Mangarrayi, Maybrat, Ndjébbana
Eurasia	5 languages: Chechen, English, Evenki, Finnish, Korean
North America	5 languages: Chinook (Lower), Diegueño (Mesa Grande), Halkomelem, Lacandón, Takelma
SE Asia and Oceania	5 languages: Apatani, Karen (Pwo), Mandarin, Timugon, Vietnamese
South America	5 languages: Cayuvava, Epena Pedee, Kamaiurá, Páez, Yagua

and Karakalpak/Turkmen). And with ten languages from Africa, the sample is not exactly balanced geographically either.

So for good reasons typologists insist that genealogical and areal diversity should be taken into account in language sampling. Let us therefore finally look at two samples that are genealogically stratified. The sample of Table 5 has one language per genus, but otherwise allows related languages to be represented. It is also areally stratified in that it has five languages from each of Dryer's macro-areas.

Here we see another prepositional SOV language (Iraqw, a Southern Cushitic language), and with two VSO languages, one AdjN, other NAdj, no evidence for Universal 17 is left.

Finally, Table 6 shows a sample based on Rijkhoff & Bakker's (1998) sampling method and Ruhlen's (1987) classification of the world's languages.

The problems with Ruhlen's ultra-lumping classification are well known, and when looking at the list of languages, many typologists will feel that the Americas are seriously underrepresented, while Eurasia is overrepresented. Nevertheless, this sample, too, confirms those of Greenberg's universals that do not concern VSO languages. It strikingly fails to replicate Greenberg's Universal 17, as the two VSO languages here are actually AdjN, thus confirming the doubts raised by some of the other samples.

Table 6. *A Rijkhoff & Bakker/Ruhlen-type sample*

	VSO	SVO	SOV	other
Postp	0	0	12	3
Prep	2	10	1	1
other	0	1	0	0

	NAdj	AdjN	other
DemN	4	14	0
NDem	10	1	0
other	1	0	0

	VSO	SVO	SOV	other
AdjN	2	4	8	1
NAdj	0	7	5	3
other	0	0	0	0

	NAdj	AdjN	other
NumN	5	15	0
NNum	10	0	0
other	0	0	0

Sample languages:

Africa	6 languages: Iraqw, Kera, Khoekhoe, Luo, Maninka (Western), Moro
Australia-New Guinea	5 languages: Arapesh, Gunbalang, Maybrat, Nabak, Ungarinjin
Eurasia	8 languages: Albanian, Chukchi, Evenki, Ket, Lezgian, Mundari, Telugu, Yukaghir (Tundra)
North America	3 languages: Greenlandic (West), Slave, Squamish
SE Asia and Oceania	4 languages: Cantonese, Chamorro, Dong, Hmong Njua
South America	4 languages: Awa Pit, Kariri, Macushi, Ndyuka

So what can we conclude from this little exercise? Most importantly, it illustrates what replication in typology would look like, showing that it is not only possible in principle, but prospects for replication are actually quite good. Greenberg (1963) used neither sophisticated sampling techniques nor thorough in-depth analyes of all the languages concerned. And yet most of his results are replicable and have been replicated. Typologists have realized that they should pay more attention to sampling issues than Greenberg did, but what these simulated replications suggest is that the impact of these different sampling methods has perhaps been overestimated, at least for correlations between features.

Widmann & Bakker (2006) claim that their study of small numerals across samples shows that sampling does matter, but it is important to note that they are not looking at correlations, but at proportions of different types. To the extent that different types are not evenly distributed geographically, a geographically biased sample will not be able to predict the overall proportions of types in the world's languages (as in politics: you cannot predict voting behavior for all of Germany by just asking voters in Bavaria). It seems that correlations tend to be less determined by geography than proportions (again, as in politics: for example, the Germany-wide correlation between the properties "not owning a car" and "voting for the Green party" can probably be established also in Bavaria alone, even though there are fewer Green voters than in Hamburg or Berlin).

A reviewer asked: "Aren't there many instances of putative universals which have been found to be figments of inadequate sampling?" It is true that many such cases can be found in the pre-Greenbergian and the generative literature, where linguists have typically looked only at a handful of languages, often closely related languages. But with the world-wide samples of the post-Greenbergian era, there are very few cases where a typological generalization has been found to be wrong because of inadequate sampling (Dryer 1988 is a well-known but fairly unique case).

Received: 30 September 2004
Revised: 28 October 2005

Max-Planck-Institut für evolutionäre Anthropologie
Universität Leipzig

Correspondence address: (Haspelmath) Max-Planck-Institut für evolutionäre Anthropologie, Deutscher Platz 6, 04105 Leipzig, Germany; e-mail: haspelmath@eva.mpg.de

References

Bybee, Joan L., Revere D. Perkins, & William Pagliuca (1994). *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World.* Chicago: University of Chicago Press.

Dryer, Matthew S. (1988). Object-verb order and adjective-noun order: Dispelling a myth. *Lingua* 74: 77–109.

— (1989). Large linguistic areas and language sampling. *Studies in Language* 13: 257–292.

— (1992). The Greenbergian word order correlations. *Language* 68: 81–138.

— (2005a). Order of subject, object, and verb. In Haspelmath et al. (eds.) 2005, 330–333.

— (2005b). Order of adposition and noun phrase. In Haspelmath et al. (eds.) 2005, 346–349.

— (2005c). Order of adjective and noun. In Haspelmath et al. (eds.) 2005, 354–357.

— (2005d). Order of demonstrative and noun. In Haspelmath et al. (eds.) 2005, 358–361.

— (2005e). Order of numeral and noun. In Haspelmath et al. (eds.) 2005, 362–365.

Du Bois, John W., Lorraine E. Kumpf, & William J. Ashby (eds.) (2003). *Preferred Argument Structure: Grammar as Architecture for Function*. Amsterdam: Benjamins.

Greenberg, Joseph H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In Joseph H. Greenberg (ed.), *Universals of Grammar*, 73–113. Cambridge, Mass.: MIT Press.

Haspelmath, Martin, Matthew S. Dryer, David Gil, & Bernard Comrie (eds.) (2005). *The World Atlas of Language Structures.* Oxford: Oxford University Press.

Hawkins, John A. (1983). *Word Order Universals.* New York: Academic Press.

Newmeyer, Frederick J. (1998). *Language Form and Language Function.* Cambridge, Mass.: MIT Press.

Rijkhoff, Jan & Dik Bakker (1998). Language sampling. *Linguistic Typology* 2: 263–314.

Ruhlen, Merritt (1987). *A Guide to the World's Languages, Volume 1: Classification*. Stanford, Cal.: Stanford University Press.

Widmann, Thomas & Peter Bakker (2006). Does sampling matter? A test in replicability. *Linguistic Typology* 10: 83–95.